Petroleum 2 (2016) 301–306

Contents lists available at ScienceDirect

Petroleum

journal homepage: www.keaipublishing.com/en/journals/petlm

Original article

Prediction of asphaltene precipitation using support vector regression tuned with genetic algorithms

Mohammad Ghorbani [a], Ghasem Zargar [a, *], Hooshang Jazayeri-Rad [b]

[a] Department of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran
[b] Department of Instrumentation and Automation, Petroleum University of Technology (PUT), Ahwaz, Iran

ARTICLE INFO

Article history:
Received 15 November 2015
Received in revised form
13 May 2016
Accepted 17 May 2016

Keywords:
Asphaltene precipitation
Prediction
Support Vector Regression (SVR)
Genetic Algorithm (GA)
Parameter optimization

ABSTRACT

Due to the severe and costly problems caused by asphaltene precipitation in petroleum industry, developing a quick and accurate model, to predict the asphaltene precipitation under different conditions, seems crucial. In this study, a new model, namely genetic algorithm – support vector regression (GA-SVR) is proposed, which is applied to predict the amount of asphaltene precipitation. GA is used to select the best optimal values of SVR parameters and kernel parameter, simultaneously, to increase the generalization performance of the SVR. The GA-SVR model is trained and tested on the experimental data sets reported in literature. The performance of the GA-SVR model is compared with two scaling equation models, using statistical error measures and graphical analyses. The results show that the prediction performance of the proposed model, is highly reliable and satisfactory.

1. Introduction

According to the SARA separation technique, Asphaltenes constitute the most aromatic and polar portion of petroleum. Other fractions are Saturates, Aromatics and Resins which have higher molecular weight and aromatic content respectively [1–3]. In general, asphaltenes are defined as a portion of crude oil that can be solved in some aromatic solvents such as toluene, benzene, xylene and pyridine but are insoluble in normal alkane solvents like n-pentane, n-heptane and n-decane [4].

Asphaltene precipitates from crude oil as a result of changes in oil composition, pressure and temperature, however the latter has fewer effect relative to the others. Miscible flooding processes like carbon dioxide and natural gas injection, and microbial enhanced oil recoveries can lead to asphaltene precipitation in petroleum reservoirs by modifying aforementioned parameters [5–7]. This precipitation also may occur in production and refinery facilities [8,9]. Severe problems like wettability alteration, relative permeability reduction in reservoirs and damage and blockage of flow in boreholes and surface pipelines can be results of asphaltene precipitation [10].

So, many researches and studies have been conducted to predict the amount of asphaltene precipitation in both theoretical and experimental approaches. In general, models of asphaltene precipitation can be divide into four categories: 1. Molecular thermodynamic models, which are an evolved type of two statistical and continuous thermodynamic models [11]. The basis of this type is assumption of polymeric structure of asphaltene molecules [12,13]. 2. Colloidal models, which are developed from the previous models, suppose that solubility of asphaltene particles is due to the attachment of resin molecules to their surfaces [14,15]. 3. Scaling equation models, which are simple and user-friendly correlations and are chiefly based on the experimental studies rather than theoretical investigations [16,17]. 4. Artificial Intelligence (AI) based models, which have been developed and are widely used recently based on finding the relation between the value of asphaltene precipitation and

* Corresponding author.
E-mail addresses: ghorbani.u@live.com (M. Ghorbani), zargar@put.ac.ir (G. Zargar).

Peer review under responsibility of Southwest Petroleum University.

Production and Hosting by Elsevier on behalf of KeAi

http://dx.doi.org/10.1016/j.petlm.2016.05.006

effective parameters without regarding the properties of the chemical process [18,19].

Due to the nature of asphaltene and current vagueness in its phase behavior and also because of diversity of effective parameters causing precipitation, a perfect and robust model cannot be introduced. This problem motivate us to introduce a model without the requirement of a thorough knowledge of nature of asphaltene. The proposed model is just based on experimental data and has a simple and fast applying procedure.

2. Development of the GA-SVR model

2.1. Data acquisition

To obtain a reliable and useful model, generality and accuracy of the method of gathering dataset are essential. So the data which are used in this study, are tried to be verified and accurate [20]. Two types of data have been implemented to develop and train the model. The first dataset, are chosen from Ashoori et al. work [18]. The selected crude oil is an asphaltenic crude oil with a specific gravity of 0.934. The amount of asphaltene precipitation is determined by gravimetric method. Three precipitants (pentane, hexane and heptane) with various dilution ratios at three 30, 50 and 70 °C temperatures are used. All of the experiments are done at the atmosphere pressure [18].

Second dataset are chosen from the study which is performed by Hu and Guo [21]. The oil, under study is the Caoqiao crude oil from an Oil Field in China. Seven n-alkanes (pentane, hexane, heptane, octane, nonane, decane, and dodecane) are used as precipitants in four temperatures including, 19.85, 34.85, 49.85 and 64.85 °C [21].

The input datasets of the proposed model include dilution ratio (Injected n-alkane volume to weight of crude oil), molecular weight of n-alkane and temperature while the corresponding amount of asphaltene precipitation (weight percent) has been considered as the target data of the model. Ranges and average of parameters of both datasets are summarized in Table 1.

2.2. Basic idea of Support Vector Regression

Support Vector Machine (SVM) is a machine learning technique based on the statistical learning theory which is proposed by Vapnic et al. in 1995 [22]. Originally, SVM developed for solving the classification problems but latter, SVR evolved from the SVM for doing regression tasks. Therefore SVM is a general term which can be divided into two subgroups; Support Vector Classification (SVC) and Support Vector Regression. In this study, the latter will be used. Assume that there is a learning samples set, $D = \{(x_i, y_i)\}$ where $x_i \in \mathbb{R}^m$ represent the input values and $y_i \in \mathbb{R}$ are the corresponding output values for $i = 1, 2, \ldots, N$ where N is the number of the samples in the training dataset and m is the dimension of the input dataset. The SVR function is generally defined as follows:

$$f(x) = \langle w, \Phi(x) \rangle + b \tag{1}$$

where $w \in \mathbb{R}^m$ is the weight vector and $b \in \mathbb{R}$ is the threshold and Φ is the mapping function which transfers the input values from a $\mathbb{R}^m$ space to a feature space with higher dimension. To train the SVR, the values of w and b must be found which are done by minimizing the following regularized risk function:

$$R_{reg}(f) = \frac{1}{2}\|w\|^2 + \frac{C}{N}\sum_{i=1}^{N} L(f(x_i), y_i) \tag{2}$$

where L is the error loss function. The term $\frac{1}{N}\sum_{i=1}^{N} L(f(x_i), y_i)$ is the average loss over the training samples, i.e. it represents the empirical risk, and $\frac{1}{2}\|w\|^2$ is the regularization term. C is the error penalty parameter which defines a trade-off between an approximation error and the weights vector norm $\|w\|$, and is chosen by the user.

According to the structural risk minimization principle, minimization of the above regularized risk function leads to the following Quadratic Programming (QP) problem:

$$\min_{w,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{3}$$

subject to:

$$y_i - \langle w, \Phi(x_i) \rangle - b \le \varepsilon + \xi, \quad i = 1, \ldots, N$$

$$\langle w, \Phi(x_i) \rangle + b - y_i \le \varepsilon + \xi^*, \quad i = 1, \ldots, N$$

$$\xi \ge 0, \quad i = 1, \ldots, N$$

$$\xi^* \ge 0, \quad i = 1, \ldots, N$$

where ξ and ξ^* are slack variables.

After converting the above problem to a dual Lagrangian problem and solving the dual problem, the regression function is written as:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\langle \Phi(x_i), \Phi(x) \rangle + b \tag{4}$$

The dot product in the above equation, can be replaced by kernel function $k(x_i, x)$, which is a useful trick to avoid the difficulties encountered with nonlinear mapping into a higher dimension space. This replacement allows to restate the support vector regression function as:

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{5}$$

There are various types of kernel functions, from which more common ones are listed in Table 2. Because of the acceptable

Table 1
The ranges of the data are used in the GA-SVR model.

Type	Parameter	Dataset					
		Ashoori et al. work			Hu and Guo work		
		Min	Max	Ave	Min	Max	Ave
Inputs	Dilution ratio, mL/g	0.67	20	7.62	2.3	24.3	12.08
	Temperature, °C	30	70	50	19.85	64.85	39.28
	Molecular weight	72.15	100.2	86.17	72.15	170.33	116.14
Output	Asphaltene precipitation, %	0.5	10.4	4.78	0.12	7.06	2.96

Table 2
Kernel functions.

Kernel function	Formula
Polynomial	$k(x_i, x) = (\langle x_i, x \rangle + 1)^d$
Gaussian Radial Basis	$k(x_i, x) = exp\left(-\frac{\|x_i - x\|^2}{2\gamma^2}\right)$
Sigmoid	$k(x_i, x) = \tanh(\rho\langle x_i, x \rangle + \varrho)$

generalization ability of the Gaussian radial basis function (RBF) kernel, this kernel is used in this study as the kernel function.

2.3. Optimization of the SVR parameters

There are two hyper-parameters in SVR formulation, C and ε and one parameter in kernel function, γ in RBF kernel. The effectiveness and generalization of the model highly depend on the accurate selection of these three parameters [23]. As mentioned earlier, C is the penalty parameter which specifies the trade-off between the empirical risk and the model complexity. Another SVR hyper-parameter is ε, which defines the size of the ε-tube around the regression function which influences the number of support vectors [24]. RBF kernel parameter γ, is the variance of kernel function which determines the nonlinear mapping of the input data to the feature space. All these parameters, are user-defined parameters. Various methods are proposed for selecting these parameters until now [25–28]. It can be said that the most popular method is the cross-validation method. In this study GA is applied to select the optimal values of the SVR parameters.

2.4. Genetic algorithm for parameter selection

Genetic Algorithm is a stochastic search method based on the principle of the survival of the fittest, gained attention after the work of John Holland in the early 1970s [29]. GA belongs to Evolutionary Algorithm (EA) family which tries to solve the optimization problems by using techniques which are based on natural evolution. The algorithm is well suited for finding the global optimal solution in a complicated multidimensional search space [30]. In this study, GA is used to select the optimal parameters of the SVR model simultaneously to reach a better generalization performance of the model.

The general procedure of GA for selecting optimal parameters is shown in Fig. 1. Steps are as follows:

1. Generating the initial population randomly.
2. Evaluating each individual in the population using a defined fitness function.
3. Bringing out the offspring by selection, crossover and mutation operators.
4. Inspecting the stopping conditions;
   i) If the end condition is met, choose the individual as the optimal parameters set
   ii) Otherwise, repeat the steps from step 2 on the new generated population.

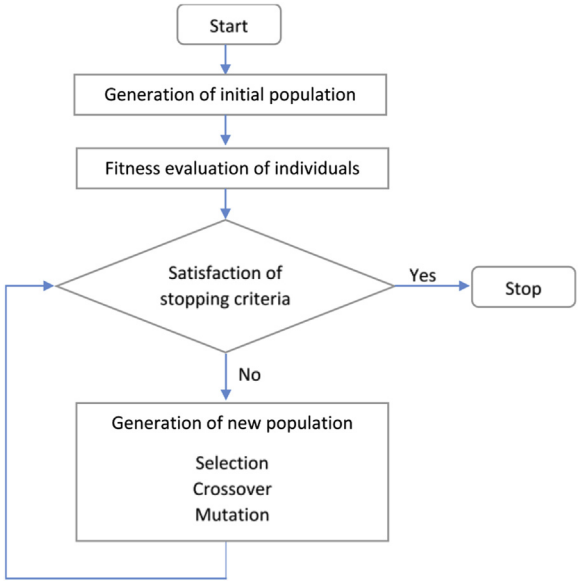

Fig. 1. Flow chart of the Genetic Algorithm.

In the present study, the initial population is generated randomly with the population size of 30. This size selected after examination of various population sizes. The individuals of the population consist of the SVR hyper-parameters, C and ε and RBF kernel function shape parameter (γ).

Mean 5-fold cross validated normalized mean square error is used as the fitness function with following equation:

$$Fittness\ Function = \frac{1}{5}\sum_{k=1}^{5} NMSE_k \tag{6}$$

where

$$NMSE = \frac{\sum_{i=1}^{N}(d_i - p_i)^2}{\sum_{i=1}^{N}\left(d_i - \overline{d}_i\right)^2} \tag{7}$$

k is the number of cross validation fold. d_i, p_i and $\overline{d}_i$ denote the desired (actual) value, the predicted value and the mean of the desired values, respectively. N is the total number of data samples in the validation dataset for each fold.

New population is created using roulette wheel method as the selection operator, heuristic method with ratio set to 1.2 as the crossover operator and adaptive feasible method as the mutation operator.

Roulette wheel selection is clearly inspired from a roulette wheel in which, parent individuals are selected so that the area of the section of the wheel corresponding to an individual is proportionate to the fitness value of the individual. A random number is used to select one of the sections with a probability equal to its area.

Heuristic crossover method returns an offspring that lies on the line containing the two parent individuals with a small distance away from the parent with the better fitness value in the direction away from the parent with the worse fitness value by using the following equation:

$$child = parent2 + R*(parent1 - parent2) \tag{8}$$

where R is the parameter ratio which specifies how far the child is from the better parent. *parent*1 has the better fitness value than *parent*2.

Adaptive feasible mutation method generates directions randomly that are adaptive relating to the last successful or unsuccessful generation. The mutation chooses a direction and step length that satisfies bounds and linear constraints [31].

The stopping condition is specified using the number of generations which is set to 50, after observation of no improvement in results for further generations.

2.5. The proposed GA-SVR model

In this study, the SVM toolbox is combined with Global Optimization Toolbox of MATLAB to develop the model [31,32]. All codes are written in MATLAB environment. Radial basis function is used as the kernel function. To train the model, experimental datasets are randomly divided into two subsets,

Table 3
Statistical parameters.

Parameter	Formula
Root Mean Square Error	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - p_i)^2}$
Coefficient of Determination	$R^2 = 1 - \frac{\sum_{i=1}^{N}(d_i - p_i)^2}{\sum_{i=1}^{N}(p_i - \overline{d})^2}$
Relative Absolute Deviation	$\sigma = \sum_{i=1}^{N}\left\lvert\frac{d_i - p_i}{d_i}\right\rvert$

N is the total number of data samples.
$\overline{d}$ is the mean of the measured values.

one for training and other is used for testing the model. The training subset is used to obtain the SVR hyper-parameters and RBF width parameter. The testing data are utilized to investigate the prediction performance of the model. 80% of the all main dataset are used as the training subset and 20% as the testing subset. Indeed, this random separation was carried out several times to avoid the local accumulation of the data points and get a homogeneous one.

At first, the training inputs and outputs, range of the SVR and RBF parameters, population size and number of generations are fed to the GA. Then the initial population is generated in GA. As a result of using 5-fold cross validation for obtaining the fitness values, the original training dataset are temporarily divided into two subsets, five times, according to cross validation rules. At each time, one subset is used to train the model and the other subset is used to calculate the fitness value. Finally, a mean fitness value is obtained for each individual of the population using equation 6.

This procedure is repeated until the all individuals be evaluated. After that, a new population is generated using genetic operators, i.e. selection, crossover and mutation operators. Generation of the new Populations continues for 50 times. Then, the best optimal parameters are chosen using the individual with the best fitness value at the last generation.

At the final step, the model is used to predict the testing outputs. The results are compared through statistical and graphical approaches. Note that the final assessment of the model performance is based on the testing data.

3. Result and discussion

In order to check the effectiveness of the proposed model, three statistical measures of measured (d) and predicted (p) asphaltene precipitation values are used, including Root Mean Square Error (RMSE), Coefficient of Determination (R^2) and Relative Absolute Deviation (σ), which are described in Table 3. Smaller RMSE and σ and higher R^2 of a model, are generally regarded as the better generalization and higher capability of that model in the prediction tasks.

The results of the statistical analyses for the proposed GA-SVR model with the optimal parameters are shown in Table 4. The result confirms the generalization and accuracy of the model for both testing and training datasets even for various data types. It is obvious that the performance of the models depends on the better choices of the SVR and kernel parameters. Values of the optimal parameters for the given data types, are listed in Table 5.

Table 5
Optimal parameters for the GA-SVR model.

Used data	Optimal Parameters		
	C	E	σ
Ashoori et al.	357.36	0.002238	0.1498
Hu and Guo	577.34	0.006681	0.4361

For better comprehension and visualization of the model's performance, the residuals distribution of the testing data is sketched in Fig. 2 and Fig. 3. Also, the measured values versus predicted values for the testing data are plotted in Fig. 4 and Fig. 5.

The same datasets with the exactly same subsets as the training and testing data for each data type, have been used to estimate the amount of asphaltene precipitation using the Ashoori et al. and Hu and Guo, scaling equation models [18,21]. Statistical error measures for the Ashoori et al. and Hu and Guo scaling equation models, are also included in Table 4. From this table, can be concluded that the proposed GA-SVR model has better statistical error measures in general for different data types as opposed to the two other models that perform well only for the data from which are derived. The measured values against predicted values for the testing data are plotted in Fig. 6 and Fig. 7 for each model which are applied to their own data. There is a good agreement between the predicted values and the experimentally measured ones for the GA-SVR model's results which is applied to the different data types. As can be seen, the scaling equation models only perform well when are applied to their driven-from data but the proposed GA-SVR model can efficiently predict for different types of data which are gained by different experimental procedures.

All the above observations, confirm that the GA-SVR model, has an excellent generalization performance and can be utilized as a reliable and fast method for predicting the amount of asphaltene precipitation.

4. Conclusions

In this paper, the application of a new proposed GA-SVR model in prediction of asphaltene precipitation has been presented. Genetic algorithm is used to optimize the parameters of the SVR simultaneously to enhance the generalization performance of the model. The developed model applied to two different datasets, and in both cases, a good agreement between the predicted amount of asphaltene precipitations and measured ones has been obtained. A comparison between the proposed

Table 4
Statistical error measures for different models.

Model	Used data	Train data				Test data			
		Numbers	RMSE	R^2	σ	Numbers	RMSE	R^2	σ
GA-SVR Model	Ashoori et al.	72	0.222504	0.994458	2.929102	18	0.198982	0.994996	0.929236
	Hu and Guo	141	0.109401	0.995726	8.095307	35	0.127253	0.994262	1.540331
Ashoori et al. scaling equation	Ashoori et al.	72	0.388995	0.981675	7.765961	18	0.451989	0.970734	2.123418
	Hu and Guo	141	0.226509	0.982066	23.429536	35	0.249381	0.978728	2.862291
Hu and Guo scaling equation	Ashoori et al.	72	0.426940	0.979868	12.266543	18	0.622855	0.952120	3.164244
	Hu and Guo	141	0.145187	0.992437	12.172973	35	0.118443	0.995122	1.645053

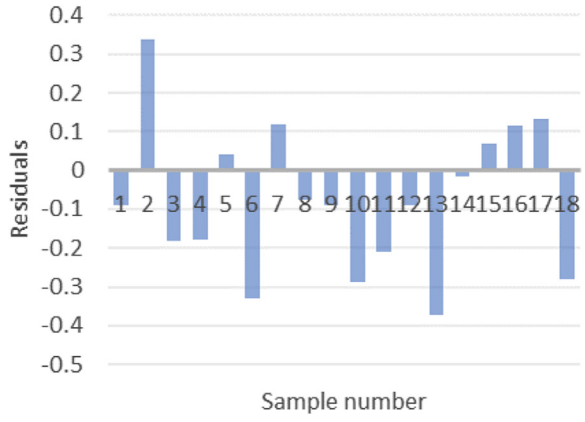

Fig. 2. Residuals of the GA-SVR model for the testing data applied to the Ashoori et al. data.

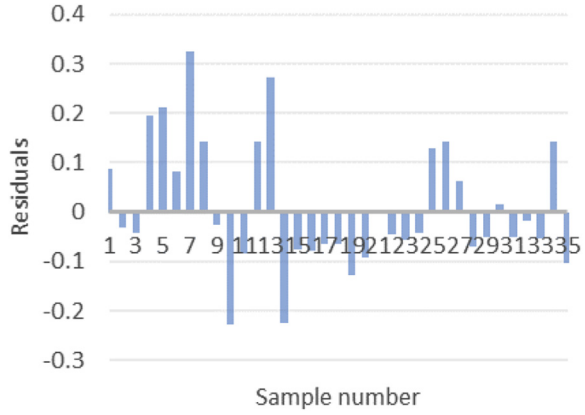

Fig. 3. Residuals of the GA-SVR model for the testing data applied to the Hu and Guo data.

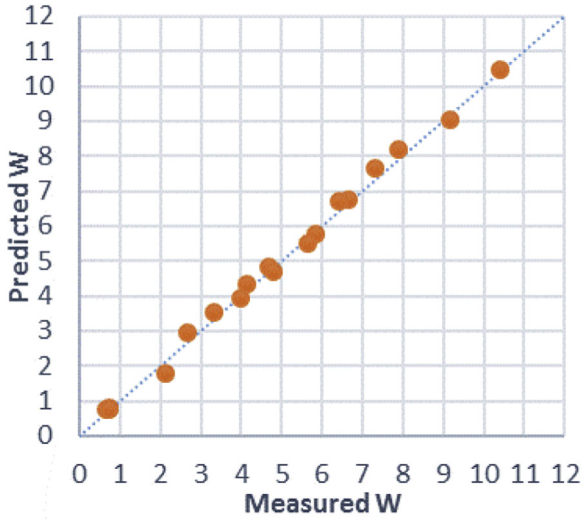

Fig. 4. Comparison of the predicted and measured values of the asphaltene precipitation for the testing data in the GA-SVR model applied to the Ashoori et al. data.

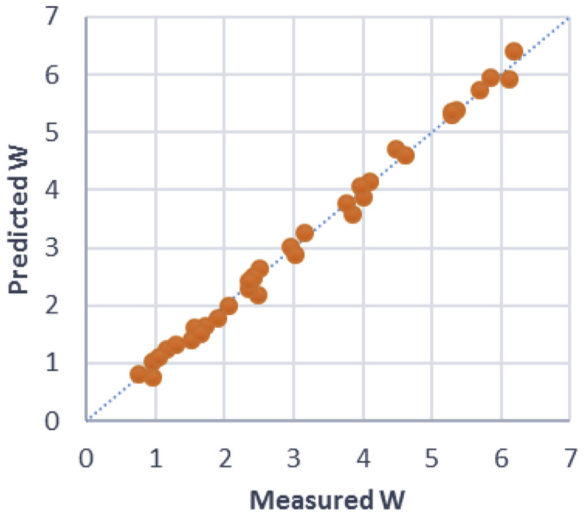

Fig. 5. Comparison of the predicted and measured values of the asphaltene precipitation for the testing data in the GA-SVR model applied to the Hu and Guo data.

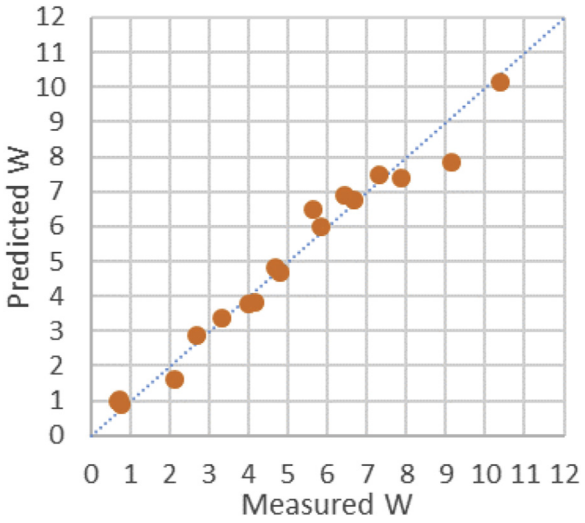

Fig. 6. Comparison of the predicted and measured values of the asphaltene precipitation by the Ashoori et al. scaling equation model.

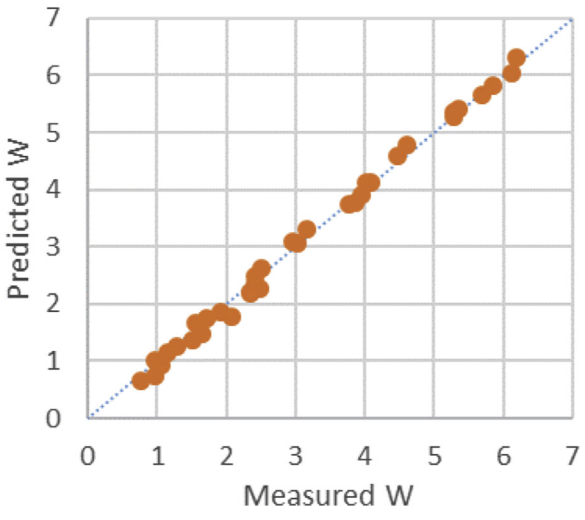

Fig. 7. Comparison of the predicted and measured values of the asphaltene precipitation by the Hu and Guo scaling equation model.

model and the two scaling equation models showed that our model performs better for different data types. The proposed model performs well at both cases regardless of the method of the experiments, so this model can replace the scaling equation models in prediction of the amount of asphaltene precipitations. Thus, the proposed model can be utilized as an accurate and fast tool in its applicability domain.

References

[1] T. Fan, J.S. Buckley, Rapid and accurate SARA analysis of medium gravity crude oils, Energy & Fuels 16 (6) (2002) 1571–1575.
[2] O.C. Mullins, E.Y. Sheu, Structures and Dynamics of Asphaltenes, Springer, 1999.
[3] K.J. Leontaritis, G. Ali Mansoori, Asphaltene deposition: a survey of field experiences and research approaches, J. Petrol. Sci. Eng. 1 (3) (1988) 229–239.
[4] L.H. Ali, K.A. Al-Ghannam, Investigations into asphaltenes in heavy crude oils. I. Effect of temperature on precipitation by alkane solvents, Fuel 60 (11) (1981) 1043–1046.
[5] H. Van Poollen, Fundamentals of Enhanced Oil Recovery, 1980.
[6] M. Blunt, F.J. Fayers, F.M. Orr Jr., Carbon dioxide in enhanced oil recovery, Energy Convers. Manag. 34 (9) (1993) 1197–1204.
[7] R.N. Mehta, V.R. Jain, Microbial enhanced oil recovery, in: Separation Processes: Emerging Technologies for Sustainable Development, 434, 2009.
[8] R. Thawer, D.C. Nicoll, G. Dick, Asphaltene deposition in production facilities, SPE Prod. Eng. 5 (04) (1990) 475–480.
[9] Z.M. Liu, G.Y. Yang, Y. Lu, B.X. Han, H.K. Yan, Phase equilibria of the CO2–Jiangsu crude oil system and precipitation of heavy components induced by supercritical CO2, J. Supercrit. Fluids 16 (1) (9/15/1999) 27–31.
[10] K. Leontaritis, J. Amaefule, R. Charles, A systematic approach for the prevention and treatment of formation damage caused by asphaltene deposition, SPE Prod. Facil. 9 (03) (1994) 157–164.
[11] A.M. Mofidi, M. Edalat, A simplified thermodynamic modeling procedure for predicting asphaltene precipitation, Fuel 85 (17) (2006) 2616–2621.
[12] A. Hirschberg, L. DeJong, B. Schipper, J. Meijer, Influence of temperature and pressure on asphaltene flocculation, Soc. Petrol. Eng. J. 24 (03) (1984) 283–293.
[13] L.X. Nghiem, D.A. Coombe, S. Farouq Ali, Compositional simulation of asphaltene deposition and plugging, in: SPE Annual Technical Conference, 1998, pp. 129–140.
[14] K. Leontaritis, G. Mansoori, Asphaltene flocculation during oil production and processing: a thermodynamic colloidal model, in: Paper SPE, 16258, 1987, pp. 4–6.
[15] H. Pan, A. Firoozabadi, A thermodynamic micellization model for asphaltene precipitation: Part I: Micellar size and growth, in: SPE Annual Technical Conference, 1996, pp. 593–605.
[16] H. Rassamdana, B. Dabir, M. Nematy, M. Farhani, M. Sahimi, Asphalt flocculation and deposition: I. The onset of precipitation, AIChE J. 42 (1) (1996) 10–22.
[17] A.K. Menshad, A.M. Mofidi, F. Shariatpanahi, M. Edalat, Developing of scaling equation with function of pressure to determine onset of asphaltene precipitation, J. Jpn. Petrol. Inst. 51 (2) (2008) 102–106.
[18] S. Ashoori, A. Abedini, R. Abedini, K.Q. Nasheghi, Comparison of scaling equation with neural network model for prediction of asphaltene precipitation, J. Petrol. Sci. Eng. 72 (1) (2010) 186–194.
[19] M.A. Ahmadi, Prediction of asphaltene precipitation using artificial neural network optimized by imperialist competitive algorithm, J. Petrol. Explor. Prod. Technol. 1 (2–4) (2011) 99–106.
[20] A.H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, A novel method for evaluation of asphaltene precipitation titration data, Chem. Eng. Sci. 78 (2012) 181–185.
[21] Y.F. Hu, T.M. Guo, Effect of temperature and molecular weight of n-alkane precipitants on asphaltene precipitation, Fluid Phase Equilibria 192 (1) (2001) 13–25.
[22] V. Vapnik, The Nature of Statistical Learning Theory, springer, 2000.
[23] K. Duan, S.S. Keerthi, A.N. Poo, Evaluation of simple performance measures for tuning SVM hyperparameters, Neurocomputing 51 (2003) 41–59.
[24] V. Kecman, Support Vector Machines: Theory and Applications, Springer, 2005, pp. 1–47.
[25] C. Campbell, N. Cristianini, J. Shawe-Taylor, Dynamically adapting kernels in support vector machines, Adv. Neural Inf. Process. Syst. 11 (1999) 204–210.
[26] Y. Zhu, C. Li, Y. Zhang, Advances in Neural Networks–ISNN 2004, Springer, 2004, pp. 518–523.
[27] P. Rivas-Perea, J. Cota-Ruiz, J.P. Venzor, D.G. Chaparro, J.G. Rosiles, LP-SVR model selection using an inexact globalized Quasi-Newton Strategy, J. Intelligent Learn. Syst. Appl. 5 (2013) 19.
[28] J.E. Wang, J.Z. Qiao, Parameter selection of SVR based on improved K-Fold Cross validation, Appl. Mech. Mater. 462 (2014) 182–186.
[29] J.H. Holland, Adaptation in Natural and Artificial Systems: an Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence, U Michigan Press, 1975.
[30] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, 1989.
[31] Global Optimization Toolbox User's Guide, The MathWorks, Inc, 2014.
[32] Canu, S., Grandvalet, Y., Guigue, V., and Rakotomamonjy, A. SVM and Kernel Methods Matlab Toolbox. 2005.